

Weekly Report

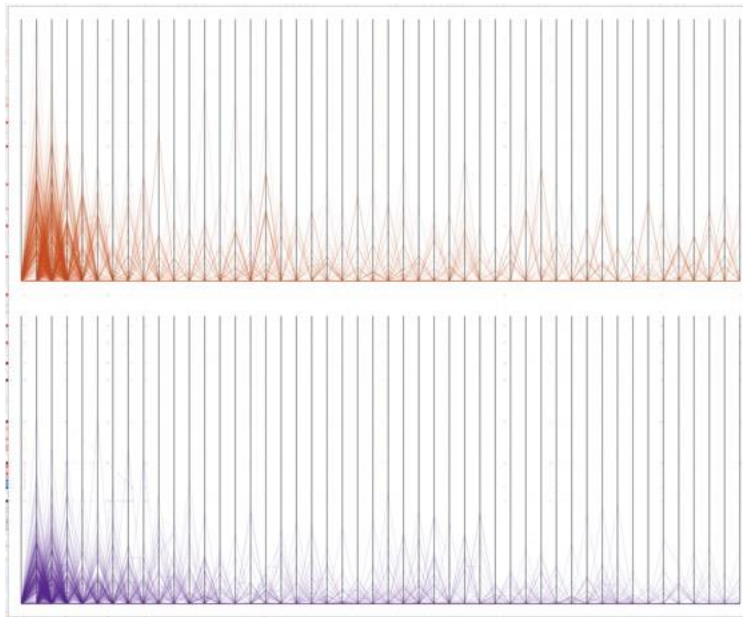
1 Done

1.1 MOOC Courses

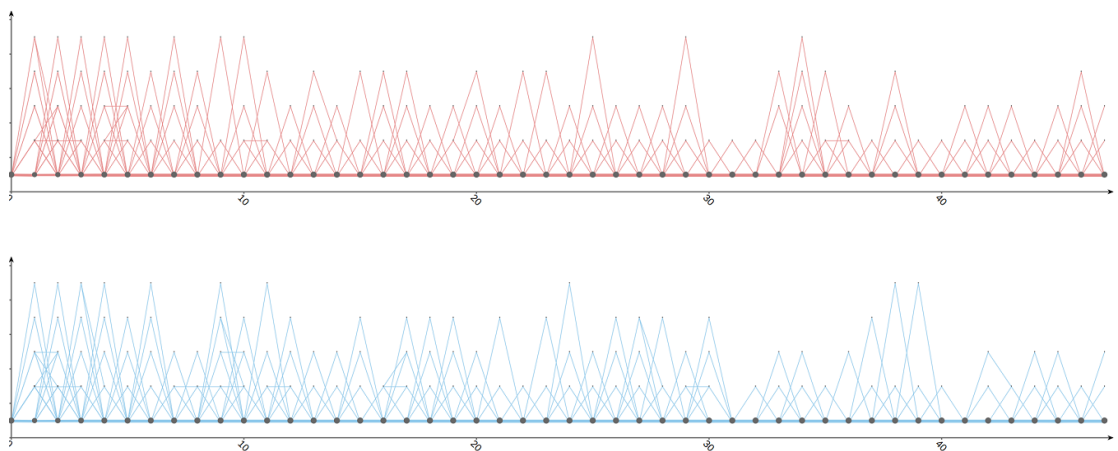
- Design exam questions for MOOC.

1.2 Project with Chuan

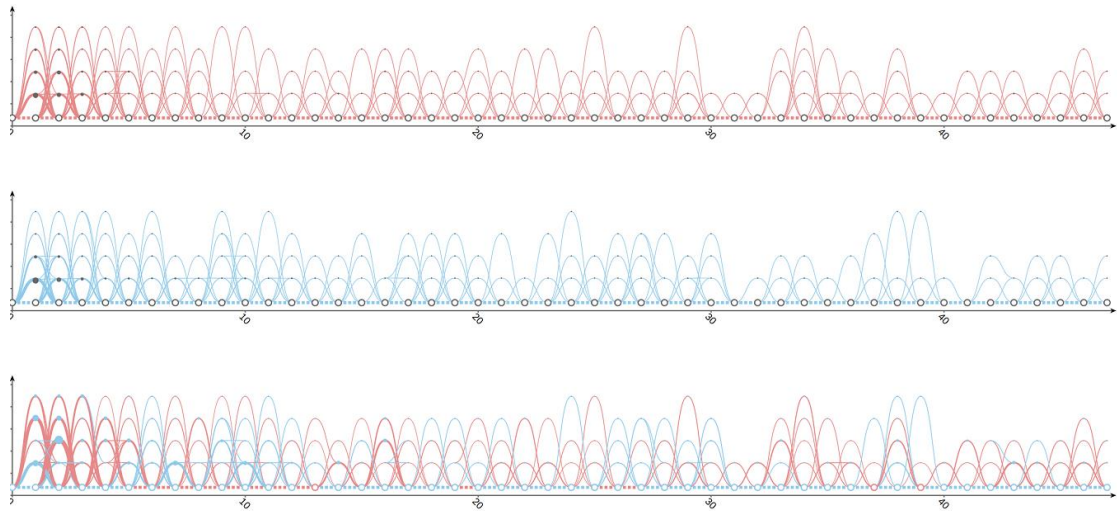
Tried some representations for intermediate result of the model. In the first time, we drew all lines in two classes, respectively. There are too many lines overlapped. So, the differences are not obvious. Besides, most values are null (default as zero), which leads to bad clustering effects.



Then, we removed the records without non-zero values. However, most values are still zero. So, we binned the values and clustered lines.



In order to make the chart beautiful, we use curves instead of straight lines. Moreover, the third chart summarizes the differences.



1.3 Paper Reading

- *Data Utility Maximization When Leveraging Crowdsensing in Machine Learning*

To exam utility, they consider uncertainty and weighted density. By uncertainty, they choose the instances that hardest (or most ambiguous) to the learning model. We can measure uncertainty of an instance by confidence-based, margin-based and entropy-based measures. But choosing the instance hardest to the learning model might not always be the best strategy. The chosen instance may be a outlier. An alternative strategy would be to collect an instance representative of uncertain data instance.

- *Learning to select data for transfer learning with Bayesian*

Optimization

This work shows how to select n data records with highest metric values to generate best models with least data records.

They utilize two metrics:

- 1) Similarity: domain to target
- 2) Diversity: number of types, entropy, etc.

- *Class Subset Selection for Transfer Learning using Submodularity*

The main algorithm is:

1. Initialize $S = \emptyset$.
2. Let $s = \arg \max_{s' \in C} F(S \cup \{s'\}) - F(S)$ such that $|S \cup \{s'\}| \leq k$.
3. If $s \neq \emptyset$ then $S = S \cup \{s\}$ and go to step 2.
4. S is the required subset.

, where F is a set function.